1
Introduction

1.1
An Unanswered Question

The aspects of music explored in this book—meter, phrase structure, contrapuntal structure, pitch spelling, harmony, and key—are well known and, in some ways, well understood. Every music student is taught to label chords, to spell notes correctly, to identify modulations, to identify a piece as being in 3/4 or 4/4, and to recognize the phrases of a sonata and the voices of a fugue. At more advanced levels of musical discourse, these structures are most often simply taken for granted as musical facts. It is rarely considered a contribution to music theory to identify the phrases or the sequence of harmonies in a piece, nor is there often disagreement about such matters. In psychology, too, each of these facets of music has been explored to some extent (some to a very considerable extent), and there are grounds for believing that all of them are important aspects of music cognition, not merely among trained musicians but among listeners in general.

In short, there appears to be broad agreement as to the general character of these structures, the particular form they take in individual pieces, and their reality and importance in music cognition. In another respect, however, our knowledge of these aspects of music is much less advanced. If we assume that harmony, metrical structure, and the like are real and important factors in musical listening, then listening must involve extracting this information from the incoming notes. How, then, is this done; by what process are these structures inferred? At present, this is very much an open question. It is fair to say that no fully satisfactory answer has been offered for any of the kinds of structure listed above; in some areas, answers have hardly even been proposed. I will present a general approach to this problem, based on the concept of *preference rules*, which leads to highly effective procedures for inferring these kinds of information from musical inputs. Because my approach is computational rather than experimental, I must be cautious in my claims about the psychological validity of the models I propose. At the very least, however, the current approach provides a promising hypothesis about the cognition of basic musical structures which warrants further consideration and study.

While exploring processes of information extraction is my main goal, the framework I propose also sheds light on a number of other issues. First of all, music unfolds in time; we do not wait until the end of a piece to begin analyzing it, but rather, we interpret it as we go along, sometimes revising our interpretation of one part in light of what happens afterwards. Preference rule systems provide a useful framework for characterizing this real-time process. The preference rule approach also provides insight into other important aspects of musical experience, such as ambiguity, tension, and expectation. Finally, as well as providing a powerful theory of music perception, the preference rule approach also sheds valuable light on what are sometimes called the "generative" processes of music: composition and performance. I will argue that preference rule systems play an important role in composition, acting as fundamental—though flexible—constraints on the compositional process. In this way, preference rules can contribute not only to the description of music perception, but of music itself, whether at the level of musical styles, individual pieces, or structural details within pieces. The preference rule approach also relates in interesting ways to issues of musical performance, such as performance errors and expressive timing.

An important question to ask of any music theory is what corpus of music it purports to describe. My main concern in this book is with Western art music of the eighteenth and nineteenth centuries: what is sometimes called "common-practice" music or simply "tonal" music.[1] I have several reasons for focusing on this corpus. First, this is the music with which I have the greatest familiarity, and thus the music about which I am most qualified to theorize. Second, common-practice music brings with it a body of theoretical and experimental research which is unparalleled in scope and sophistication; the current study builds on this earlier work in many ways which I will do my best to acknowledge. Third, a large amount of music from the common-practice corpus is available in music notation. Music notation provides a representation which is convenient for study and can also easily be converted into a format suitable for computer analysis. This contrasts with much popular music and non-Western music, where music notation is generally not available. (There are problems with relying on music notation as well, as I will discuss below.) Despite this limited focus, I believe that many aspects of the model I present are applicable to kinds of music outside the Western canon, and at some points in the book I will explore this possibility.

Another question arises concerning the subject matter of this study. No one could deny that the kinds of musical structure listed above are important, but music has many other important aspects too. For example, one could also cite motivic structure (the network of melodic segments in a piece that are heard as similar or related); melodic schemata such as the gap-fill archetype (Meyer 1973) and the $\hat{1}$-$\hat{7}$-$\hat{4}$-$\hat{3}$ schema (Gjerdingen 1988); and the conventional "topics"—musical gestures with extramusical meanings—discussed by Ratner (1980) and others. In view of this, one might ask why I consider only the aspects of music listed earlier. An analogy may be useful in explaining what these kinds of musical structure have in common, and the role they play in music cognition.

Any regular observer of the news media will be familiar with the term "infrastructure." As the term is commonly used, "infrastructure" refers to a network of basic structures and services in a society—largely related to transportation and communication—which are required for the society to function. (The term is most often heard in the phrase "repairing our crumbling infrastructure"—a frequent promise of politicians.) To my mind, "infrastructure" implies two important things. Infrastructure is supposed to be ubiquitous: wherever you go (ideally), you will find the roads, power lines, water mains, and so on that are needed for life and business. Secondly, infrastructure is a means to an end: water mains and power lines do not normally bring us joy in themselves, but they facilitate other things—homes, schools, showers, VCRs—whose contribution to life is more direct. In both of these respects, the aspects of music listed earlier could well be regarded as an "infrastructure" for tonal music. Metrical structure and harmony are ubiquitous: roughly speaking, every piece, in fact every moment of every piece, has a metrical structure and a harmonic structure. Melodic archetypes and topics, by contrast, are *occasional* (though certainly common). Few would argue, I think, that every bit of tonal music is a melodic archetype or a topic. Secondly, while the structures I discuss here may sometimes possess a kind of direct musical value in their own right, they function largely as means to other musical ends. In many cases, these musical ends are exactly the kinds of occasional structures just mentioned. A topic or melodic archetype requires a certain configuration of contrapuntal, metrical, and harmonic structures, and perhaps others as well; indeed, such higher-level patterns are often characterized largely in infrastructural terms (I will return to this point in chapter 12). My aim here is not, of course, to argue for either "ubiquitous" or "occasional" structures as more important than the other—each is important in its own way; my point, rather, is that ubiquitous structures form a "natural kind" and, hence, an appropriate object of exclusive study.

1.2 Goals and Methodology

Discourse about music adopts a variety of methods and pursues a variety of goals. In this section I will explain the aims of the current study and my method of achieving them. It is appropriate to begin with a discussion of the larger field in which this study can most comfortably be placed, a relatively new field known as *music cognition*.

Music cognition might best be regarded as the musical branch of cognitive science—an interdisciplinary field which has developed over the last thirty years or so, bringing together disciplines relating to cognition, such as cognitive psychology, artificial intelligence, neuroscience, and linguistics. Each of the disciplines contributing to cognitive science brings its own methodological approach; and each of these methodologies has been fruitfully applied to music. The methodology of cognitive psychology itself is primarily experimental: human subjects are given stimuli and asked to perform tasks or give verbal reports, and the psychological processes involved are inferred from these. A large body of experimental work has been done on music cognition; this work will frequently be cited below. In theoretical linguistics, by contrast, the methodology has been largely introspectionist. The reasoning in linguistics is that, while we do not have direct intuitions about the syntactic structures of sentences, we do have intuitions about whether sentences are syntactically well-formed (and perhaps about other things, such as whether two sentences are identical in meaning). These well-formedness judgments constitute a kind of data about linguistic understanding. By simply seeking to construct grammars that make the right judgments about well-formedness—linguists reason—we will uncover much else about the syntactic structure of the language we are studying (and languages in general). The introspectionist approach to music cognition is reflected in work by music theorists such as Lerdahl and Jackendoff (1983) and Narmour (1990).

(This is not to say, however, that music theory in general should be regarded as introspectionist cognitive science; I will return to this point.)

The methods of artificial intelligence are also important in music cognition. Here, attempts are made to gain insight into a cognitive process by trying to model it computationally. Often, the aim is simply to devise a computational system which can perform a particular process (for example, yielding a certain desired output for a given input); while there is no guarantee that such a program performs the process the same way humans do it, such an approach may at least shed some light on the psychological mechanisms involved.[2] In some cases, this approach has received empirical support as well, in that neurological mechanisms have been found which actually perform the kind of functions suggested by computational models (see Bruce & Green 1990, 87–104, for discussion of examples in the area of vision). As we will see, this, too, is a widely used approach in music cognition. Finally, cognition can be approached from a neurological or anatomical perspective, through studies of electric potentials, brain disorders, and the like. This approach has not been pursued as much as others in music cognition, though some progress has been made; for example, much has been learned regarding the localization of musical functions in the brain.[3]

Despite their differing methodologies, the disciplines of cognitive science share certain assumptions. All are concerned with the study of intelligent systems, in particular, the human brain. It is widely assumed, also, that cognitive processes involve representations, and that explanations of cognitive functions should be presented in these terms. This assumption is very widely held, though not universally.[4] To appreciate its centrality, one need only consider the kinds of concepts and entities that have been proposed in cognitive science: for example, edge detectors and primal sketches in vision, tree structures and constituents in linguistics, prototypes and features in categorization, networks and schemata in knowledge representation, loops and buffers in memory, problem spaces and productions in problem-solving, and so on. All of these are kinds of mental representations, proposed to explain observed facts of behavior or introspection. A second important assumption is the idea of "levels of explanation." A cognitive process might be described at a neurological level; but one might also describe it at a higher, computational level, without worrying about how it might be instantiated neurologically. A computational description is no less real than a neurological one; it is simply more abstract. It is assumed, further, that a cognitive system, described at a computational level, might be physically instantiated in quite different ways: for example, in a human brain or on a computer. This assumption is crucial for artificial intelligence, for it implies that a computer running a particular program might be put forth as a description or model of a cognitive system, albeit a description at a very abstract level.[5]

This background may be helpful in understanding the goals and methodology of the current study. My aim in this study is to gain insight into the processes whereby listeners infer basic kinds of structure from musical input. My concern is with what Lerdahl and Jackendoff (1983, 3) call "experienced listeners" of tonal music: people who are familiar with the style, though not necessarily having extensive formal training in it. My methodology in pursuing this goal was both introspectionist and computational. For a given kind of structure, it was first necessary to determine the correct analysis (metrical, harmonic, etc.) of many musical excerpts. Here my approach was mainly introspective; I relied largely on my own intuitions as to the correct analyses of pieces. However, I sometimes relied on other sources as well. With some of the kinds of structure explored here, the correct analysis is at least partly explicit in music notation. For example, metrical structure is indicated by rhythmic notation, time signatures, and barlines. For the most part, the structures implied by the notation of pieces concur with my own intuitions (and I think those of most other listeners), so notation simply provided added confirmation.[6] I then sought models to explain how certain musical inputs might give rise to certain analyses; and I devised computational implementations of these models, in order to test and refine them. With each kind of structure, I performed a systematic test of the model (using some source other than my own intuitions for the correct analysis—either the score or analyses done by other theorists) to determine its level of success.

The goals and methodology I have outlined could be questioned in several ways. The first concerns the computational nature of the study. As mentioned earlier, the mere fact that a model performs a process successfully certainly does not prove that the process is being performed cognitively in the same way. However, if a model does *not* perform a process successfully, then one knows that the process is *not* performed cognitively in that way. If the model succeeds in its purpose, then one has at least a hypothesis for how the process might be performed cognitively, which can then be tested by other means. Computer implementations are also valuable, simply because they allow one to test objectively whether a model can actually produce the desired outputs. In the current case, the programs I devised often did not produce the results I expected, and led me to modify my original models significantly.

Another possible line of criticism concerns the idea of "correct" analyses, and the way I arrived at them. It might seem questionable for me, as a music theorist, to take my intuitions (or those of another music theorist) about musical structure to represent those of a larger population of "experienced listeners." Surely the hearing of music theorists has been influenced (enhanced, contaminated, or just changed) by very specialized and unusual training. This is, indeed, a problematic issue. However, two points should be borne in mind. First, it is certainly not out of the question that untrained and highly trained listeners have much in common in at least some aspects of their music cognition. This is of course the assumption in linguistics, where linguists take their own intuitions about syntactic well-formedness (despite their highly specialized training in this area) to be representative of those of the general population. Secondly, and more decisively, there is an impressive body of experimental work suggesting that, broadly speaking, the kinds of musical representations explored here *are* psychologically real for a broad population of listeners; I will refer to this work often in the chapters that follow. Still, I do not wish to claim that music theorists hear things like harmony, key, and so on exactly the same way as untrained listeners; surely they do not. Much further experimental work will be needed to determine how much, and in what ways, music cognition is affected by training.

Quite apart from effects of training, one might argue that judgments about the kinds of structures described here vary greatly among individuals—even among experts (or non-experts). Indeed, one might claim that there is so much subjectivity in these matters that the idea of pursuing a "formal theory of listeners' intuitions" is misguided.[7] I do not deny that there are sometimes subjective differences about all of the kinds of structure at issue here; however, I believe there is much more agreement than disagreement. The success of the computational tests I present here, where I rely on sources other than myself for the "correct" analysis, offers some testimony to the general agreement that is found in these areas. (One might also object that, even for a single listener, it is oversimplified to assume that a single analysis is always preferred to the exclusion of all others. This is certainly true; ambiguity is a very real and important part of music cognition, and one which is considerably illuminated by a preference rule approach, as I discuss in chapter 8.)

An important caveat is needed about the preceding discussion. My concern here is with aspects of music perception which I assume to be shared across a broad population of listeners familiar with tonal music. I must emphasize, however, that I am not at all assuming that these principles are innate or universal. Rather, it is quite possible that they are learned largely from exposure to music—just as language is, for example (at least, some aspects of language). I will argue in later chapters that some aspects of the models I propose have relevance to kinds of music outside the Western canon. However, I will take no position on the questions of universality and innateness; in my view, there is not yet sufficient basis for making claims about these matters.

1.3 Music Cognition and Music Theory

I suggested above that some work in music theory might be regarded as introspectionist cognitive science—work seeking to reveal cognitive processes through introspection, much as linguists do with syntax. Indeed, music theory has played an indispensable role in music cognition as a source of models and hypotheses; much music-related work in cognitive psychology has been concerned with testing these ideas. However, it would be a mistake to regard music theory in general as pursuing the same goals as music cognition. Cognitive science is concerned, ultimately, with describing and explaining cognitive processes. In the case of music cognition, this normally implies processes involved in listening, and sometimes performance; it might also involve processes involved in composition, although this area has hardly been explored. I have argued elsewhere that, while some music theory is concerned with this goal, much music theory is not; rather, it is concerned with enhancing our listening, with finding new structures in pieces which might enrich our experience of them (Temperley in press-b). Many music theorists state this goal quite explicitly. I have called the latter enterprise "suggestive theory"; this is in contrast to the enterprise of "descriptive theory," which aims to describe cognitive processes. Consider Z-related sets, a widely used concept in pitch-class set theory: two pitch-class sets are Z-related if they have the same intervallic content, but are not of the same set-type (related by transposition or inversion). I believe few theorists would claim that people hear Z-related sets (except as a result of studying set theory); rather, Z-related sets serve to enhance or enrich our hearing of certain kinds of music once we are aware of them.

The goal of studying pieces of music in order to understand them more fully, and to enrich our experience of them as much as possible, is an enormously worthwhile one. However, suggesting ways of enhancing our hearing is a goal quite different from describing our hearing. There is a good deal of confusion about this point in music theory, and it is often unclear how specific theories or analyses are to be construed. This is particularly apparent with Schenkerian analysis, a highly influential approach to the study of tonal music. While some theorists have construed Schenkerian theory in a psychological way, others have viewed it as a suggestive theory: a means of enhancing and expanding our hearing of tonal music. Of course, it is possible that a theory could be suggestive in some respects and descriptive in others. My own view is that some aspects of Schenkerian theory are highly relevant to cognition; in particular, Schenkerian analysis draws our attention to subtleties of contrapuntal structure which are often not explicit in notation. (I discuss this further in chapter 8.) With other aspects of Schenkerian theory the relationship to listening is less clear, especially the "reductive" or hierarchical aspect. But to exclude aspects of Schenkerian theory (or any other music theory) from a cognitive theory of tonal music is not at all to reject or dismiss them. Rather, it is simply to maintain that their value is not, primarily, as contributions to a theory of music cognition—a position that many Schenkerian analysts have endorsed.

The psychological, rather than suggestive, perspective of the current study cannot be emphasized too strongly, and should always be kept in mind. For example, when I speak of the "correct" analysis of a piece—as I often will—I mean the analysis that I assume listeners hear, and thus the one that my model will have to produce in order to be correct. I do not mean that the analysis is necessarily the best (most musically satisfying, informed, or coherent) one that can be found. (A similar point should be made about the term "preference rule." Preference rules are not claims about what is aesthetically preferable; they are simply statements of fact about musical perception.) I have already acknowledged that, in assuming a single analysis shared by all listeners, I am assuming a degree of uniformity that is not really present. In making this assumption, I do not in any way mean to deny the importance and interest of subjective differences; such differences are simply not my concern for the moment. I *do* maintain, however, that the differences between us, as listeners, are not so great that any attempt to describe our commonalities is misguided or hopeless.

1.4 The Input Representation

An important issue to consider with any computer model is the input representation that is used. The preference rule systems discussed here all use essentially the same input representation. This is a list of notes,

Figure 1.1
A "piano-roll" representation of the opening of the Gavotte from Bach's French Suite No. 5 in G major (generated from a performance by the author on a MIDI keyboard). The score for the excerpt is shown below.

giving the on-time, off-time (both in milliseconds) and pitch of each note—what I will refer to as a "note-list." We can also think of this as a two-dimensional representation, with pitch on one axis and time on the other; each pitch-event is represented as a line segment on the plane, with the length of the line corresponding to the duration of the event. Such a representation is sometimes known as a "piano-roll," since it resembles the representations of pieces used with player pianos in the early twentieth century. Figure 1.1 shows part of the piano-roll representation for a performance of a Bach Gavotte (the score for the excerpt is shown below). Pitches in the input representation are categorized into steps of the chromatic scale; following convention, integers are used to represent pitches, with middle C = 60. In an important sense, then, the pitch axis of the "piano-roll" representation is discrete, not continous. The time axis, however, is essentially continuous; pitch-events are not quantized rhythmically in any significant way (except at the very small level of milliseconds). Other acoustic information such as timbre and amplitude is excluded from the input. (Some of the models also require additional information as input; for example, several of the models require metrical structure. I will discuss this further below.)

In assuming a "piano-roll" representation as input, I am avoiding the problem of deriving pitch information from actual sound. This problem—sometimes known as "music recognition" or "automatic transcription"—has been studied extensively, and proves to be highly complex (Moorer 1977; Foster, Schloss, & Rockmore 1982; Tanguiane 1993). The sounds of the music must be separated out from the other background sounds that are always present in any natural environment; the individual frequencies that make up the sound must be grouped together to form notes; and the notes must be correctly quantized to the right pitch categories, factoring out vibrato, bad intonation, and so on. However, this process is not our concern here; in the following chapters, the existence of an accurate piano-roll representation will simply be taken for granted.

One might wonder what evidence there is that listeners actually form piano-roll representations. Of course very few people could accurately *report* such representations; but this may be because such information is largely unconscious or not easily articulated. Most evidence for the reality of piano-roll representations is indirect, and somewhat inconclusive. For example, the fact that listeners are generally able to learn a melody from hearing it (at least if they hear it enough times), and recognize it later or reproduce it by singing, suggests that they must be extracting the necessary pitch and duration information. Another possible argument for the reality of piano-roll representations is that the kinds of higher-level structures explored here—whose general psychological reality has been quite strongly established, as I will discuss—*require* a piano-roll input in order to be derived themselves. For example, it is not obvious how one could figure out what harmonies were present in a passage without knowing what notes were present. I should point out, however, that several proposals for deriving aspects of the infrastructure—specifically harmony, contrapuntal structure, and key—assume exactly this: they assume that these kinds of structure can be extracted without first extracting pitch information. These proposals will be discussed below, and I will suggest that all of them encounter serious problems. I think a case could be made, then, that the reality of "infrastructure" levels provides strong evidence for the reality of piano-roll representations, since there is no other plausible way that infrastructure levels could be derived.

It was noted above that the input representation does not contain any quantization of events in the time dimension. This is, of course, true to the situation in actual listening. In performed music, notes are not played with perfect regularity; there is usually an *implied* regularity of durations (this will be represented by the metrical structure), but within that there are many small imperfections as well as deliberate fluctuations in timing. In the tests presented below, I often use piano-roll representations that were generated from performances on a MIDI keyboard, so that such fluctuations are preserved. (The piano-roll in figure 1.1 is an example. The imperfections in timing here can easily be seen—for example, the notes of each chord generally do not begin and end at exactly the same time.) However, one can also generate piano-roll representations from a score; if one knows the tempo of the piece, the onset and duration of each note can be precisely determined. Since pieces are never played with perfectly strict timing, using "quantized" piano-roll representations of this kind is somewhat artificial, but I will sometimes do so in the interest of simplicity and convenience.

Another aspect of the piano-roll representation which requires discussion is the exclusion of timbre and dynamics.[8] As well as being important in their own right, these musical parameters may also affect the levels of the infrastructure in certain ways. For example, dynamics affects metrical structure, in that loud notes are more likely to be heard as metrically strong; timbre affects contrapuntal structure, in that timbrally similar notes tend to stream together. Dynamics could quite easily be encoded computationally (the dynamic level of a note can be encoded as a single numerical value or series of values), and incorporating dynamics into the current models would be a logical further step. With timbre, the problem is much harder. As Bregman (1990, 92) has observed, we do not yet have a satisfactory way of representing timbre. Several multidimensional representations have been proposed, but none seem adequate to capturing the great variety and richness of timbre. Studying the effect of timbre on infrastructural levels will require a better understanding of timbre itself.

1.5 The Preference Rule Approach

The approach of the current study is based on *preference rules*. Preference rules are criteria for forming some kind of analysis of input. Many possible interpretations are considered; each rule expresses an opinion as to how well it is satisfied by a given interpretation, and these opinions are combined together to yield the preferred analysis. Perhaps the clearest antecedent for preference rules is found in the Gestalt rules of perception, proposed in the 1920s; this connection will be discussed further in chapter 3.

Preference rules *per se* were first proposed by Lerdahl and Jackendoff in their *Generative Theory of Tonal Music* (1983) (hereafter *GTTM*). Lerdahl and Jackendoff present a framework consisting of four kinds of hierarchical structure: grouping, meter, time-span reduction, and prolongational reduction. For each kind of structure, they propose a set of "well-formedness rules" which define the structures that are considered legal; they then propose preference rules for choosing the optimal analysis out of the possible ones. The model of meter I present in chapter 2 is closely related to Lerdahl and Jackendoff's model; my model of phrase structure, presented in chapter 3, has some connection to Lerdahl and Jackendoff's model of grouping. Lerdahl and Jackendoff did not propose any way of quantifying their preference rule systems, nor did they develop any implementation. The current study can be seen as an attempt to quantify and implement Lerdahl and Jackendoff's initial conception, and to expand it to other musical domains. (I will have little to say here about the third and fourth components of *GTTM*, time-span reduction and prolongational reduction. These kinds of structure are less psychologically well-established and more controversial than meter and grouping; they also relate largely to large-scale structure and relationships, which sets them apart from the aspects of music considered here.)

The preference rule approach has been subject to some criticism, largely in the context of critiques of *GTTM*. The problem most often cited is that preference rules are too vague: depending on how the rules are quantified, and the relative weights of one rule to another, a preference rule system can produce a wide range of analyses (Peel & Slawson 1984, 282, 288; Clarke 1989, 11). It is true that the preference rules of *GTTM* are somewhat vague. This does not mean that they are empty; even an informal preference rule system makes empirical claims that are subject to falsification. If a preference rule system is proposed for an aspect of structure, and one finds a situation in which the preferred analysis cannot be explained in terms of the proposed rules, then the theory is falsified, or at least incomplete. It must be said that very few music theories offer even this degree of testability. The more important point, however, is that preference rule systems also lend themselves well to rigorous formalization. If the parameters of the rules can be specified, the output of the rule system for a given input can be determined in an objective way, making the theory truly testable. This is what I attempt to do here.[9]

Another criticism that has been made of preference rule systems concerns the processing of music over time. Lerdahl and Jackendoff's stated aim in *GTTM* (1983, 3–4) is to model what they call "the final state of [a listener's] understanding" of a piece. Under their conception, preference rules serve to select the optimal analysis for a complete piece, once it has been heard in its entirety. In my initial presentation of the current model (in chapters 2 through 7), I will adopt this approach as well. This "final understanding" approach may seem problematic from a cognitive viewpoint; in reality, of course, the listening process does not work this way. However, preference rule systems also provide a natural and powerful way of modeling the moment-to-moment course of processing as it unfolds during listening. I will return to this in the next section (and at greater length in chapter 8).

One notable virtue of preference rule systems is their conceptual simplicity. With a preference rule system, the rules themselves offer a high-level description of what the system is doing: it is finding the analysis that best satisfies the rules. This is an important advantage of preference rule systems over some other models that are highly complex and do not submit easily to a concise, high-level description. (Some examples of this will be discussed in the chapters that follow.) Of course, preference rule systems require some kind of implementation, and this implementation may be highly complex. But the implementation need not be of great concern, nor does it have to be psychologically plausible; it is simply a means to the end of testing whether or not the preference rule system can work. If a preference rule system can be made to produce good computational results, it provides an elegant, substantive, high-level hypothesis about the workings of a cognitive system.

1.6 The Implementation Strategy

While I have said that details of implementation are not essential to an understanding of preference rule systems, a considerable portion of this book is in fact devoted to issues of implementation. (This includes the present section, as well as the sections of following chapters entitled "Implementation.") While these sections will, I hope, be of interest to some readers, they may be skipped without detriment to one's understanding of the rest of the book. In this section I describe a general implementation strategy which is used, in various ways, in all the preference rule models in this study.

At the broadest level, the implementation strategy used here is simple. In a given preference rule system, all possible analyses of a piece are considered. Following Lerdahl and Jackendoff, the set of "possible" analyses is defined by basic "well-formedness rules." Each preference rule then assigns a numerical score to each analysis. Normally, the analytical process involves some kind of arbitrary segmentation of the piece. Many analytical choices are possible for each segment; an analysis of the piece consists of some combination of these segment analyses. For each possible analysis of a segment, each rule assigns a score; the total score for a segment analysis sums these rule scores; the total score for the complete analysis sums the segment scores. The preferred analysis is the one that receives the highest total score.

As noted above, many of the preference rules used in these models involve numerical parameters (and there are always numerical values that must be set for determining the weight of each rule relative to the others). These parameters were mostly set by trial and error, using values that seemed to produce good results in a variety of cases. It might be possible to derive optimal values for the rules in a more systematic way, but this will not be attempted here.

One might ask why it is necessary to evaluate complete analyses of a piece; would it not be simpler to evaluate short segments in isolation? As we will see, this is not possible, because some of the preference rules require consideration of how one part of an analysis relates to another. Whether an analysis is the best one for a segment depends not just on the notes in that segment, but also on the analysis of nearby segments, which depends on the notes of those segments as well as the analysis of other segments, and so on. However, the number of possible analyses of a piece is generally huge, and grows exponentially with the length of the piece. Thus it is not actually possible to generate all well-formed analyses; a more intelligent search procedure has to be used for finding the highest-scoring one without generating them all. Various procedures are used for this purpose; these will be described in individual cases. However, one technique is of central importance in all six preference rule systems, and warrants some discussion here. This is a procedure from computer science

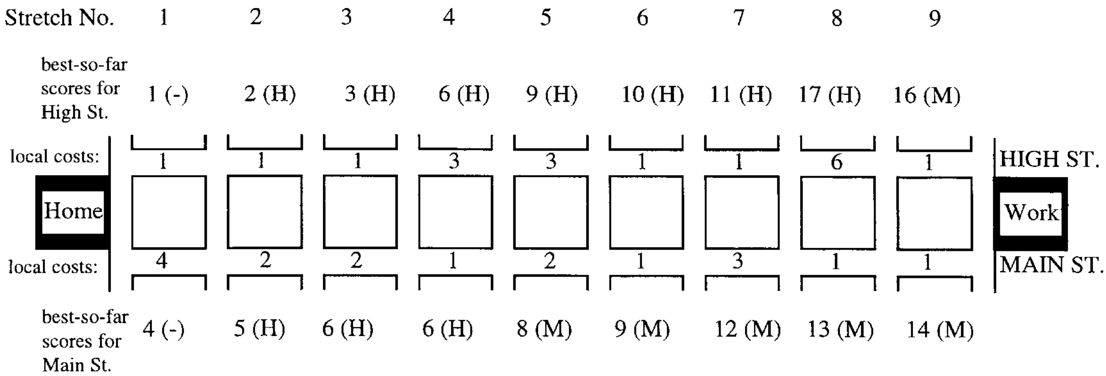

Figure 1.2

known as "dynamic programming." (The idea of using dynamic programming to implement preference rule systems is due to Daniel Sleator.)

Imagine that you are driving through a large city (see figure 1.2). You want to go from home (at the left end of the figure) to work (at the right end). There are two routes for going where you want to go; you can either drive on High Street or Main Street. The two routes are the same in total distance. However, certain stretches of each street are bad (because they have terrible potholes, or construction, or a lot of traffic). You could also switch back and forth between one street and the other at different points, but this carries a cost in terms of time. Suppose that it is worthwhile for you to really sit down and figure out the best route (perhaps because you make this trip every day). You assign each stretch of street a "cost," which is simply the number of minutes it would take you to traverse that stretch. These "local" costs are shown on each stretch of street in figure 1.2. You also assign a cost to any switch between one street and the other; say each such switch costs you 2 minutes. Now, how do you determine the best overall route? It can be seen that there are a large number of different possible routes you could take—2^n, where n is the number of blocks in the east-west direction. You could calculate the cost for every possible route; however, there is a better way. Supposing you compute the cost of all possible routes for the first two stretches that end up on High Street in stretch 2. There are only two, H-H and M-H; the best (i.e. lowest-cost) one is H-H, with a total time of 2 minutes. Then you find the best route ending up on Main Street in stretch 2; it is H-M, with a total time of 5 minutes (local costs of 1 and 2, plus a cost of 2 for switching between streets.) At this point, you do not know whether it is better to end up on High St. or Main St. in stretch 2; that depends on what happens further on. But you do know that no matter what happens later, there will never be any reason to use any route for the first two stretches other than one of the two "best-so-far" routes already identified. Now suppose we want to compute the best way of getting to Main Street in stretch 3. We can use our "best-so-far" routes to stretch 2, continuing each one in stretch 3 and calculating the new total cost; the best choice is H-H-M, with a cost of 6 minutes. Repeating the process with High Street at stretch 3, we now have two new "best-so-far" routes for stretch 3. We can continue this process all the way through to the end of the trip. At each stretch, we only need to record the best-so-far route to each ending point at that stretch, along with its score. In fact, it is not even necessary to record the entire best-so-far route; we only need to record the street that we should be on in the previous stretch. At Main Street in stretch 3, we record that it is best to be on High Street in stretch 2. In this way, each street at each stretch points back to some street at the previous stretch, allowing us to recreate the entire best-so-far route if we want to. (In figure 1.2, the score for the best-so-far route at each segment of street is shown along the top and bottom, along with the street that it points back to at the previous stretch—"H" or "M"—in parentheses.) When we get to the final stretch, either High Street or Main Street has the best (lowest) "best-so-far" score, and we can trace that back to get the best possible route for the entire trip. In this case, Main Street has the best score at the final stretch; tracing this back produces an optimal route of H-H-H-M-M-M-M-M-M.

What I have just described is a simple example of the search procedure used for the preference rule models described below. Instead of searching for the optimal path through a city, the goal is to find the optimal analysis of a piece. We can imagine a two-dimensional table, analogous to the street map in figure 1.2. Columns represent temporal segments; cells of each column represent possible analytical choices for a given segment. An analysis is a path through this table, with one step in each segment. Perhaps the simplest example is the key-finding system (described in chapter 7). Rows of the table correspond to keys, while columns correspond to measures (or some other temporal segments). At each segment, each key receives a local score indicating how compatible that key is with the pitches of the segment; there is also a "change" penalty for switching from one key to another. At each segment, for each key, we compute the best-so-far analysis ending at that key; the best-scoring analysis at the final segment can be traced back to yield the preferred analysis for the entire piece. A similar procedure is used for the harmonic analysis system (where the rows represent roots of chords, instead of keys), the pitch spelling system (where cells of a column represent possible spellings of the pitches in the segment), and the contrapuntal analysis system (where cells represent possible analyses of a segment—contrapuntal voices at different pitch levels), though there are complications in each of these cases which will be explained in due course.

The meter and phrase programs use a technique which is fundamentally similar, but also different. In the case of the phrase program, the table is simply a one-dimensional table of segments representing notes; an analysis is a subset of these notes which are chosen as phrase boundaries. (Choosing a note as a phrase boundary means that a boundary occurs immediately before that note.) Again, each note has a local score, indicating how good it is as a phrase boundary; this depends largely on the size of the temporal gap between it and the previous note. At the same time, however, it is advantageous to keep all the phrases close to a certain optimal size; a penalty is imposed for deviations from this size. At each note, we calculate the best-so-far analysis ending with a phrase boundary at that note. We can do this by continuing all the previous best-so-far analyses—the best-so-far analyses with phrase boundaries at each previous note—adding on a phrase ending at the current note, calculating the new score, and choosing the highest-scoring one to find the new best-so-far analysis. Again, we record the previous note that the best-so-far analysis points back to as well as the total score. After the final note, we compute a final "best-so-far" analysis (since there has to be a phrase boundary at the end of the piece) which yields the best analysis overall. The meter program uses a somewhat more complex version of this approach. The essential difference between this procedure and the one described earlier is that, in this case, an analysis only steps in certain segments, whereas in the previous case each analysis stepped in every segment.

Return to the city example again. Supposing the map in figure 1.2, with the costs for each stretch, was being revealed to us one stretch at a time; at each stretch we had to calculate the costs and best-so-far routes. Consider stretch 7; at this stretch, it seems advantageous to be on High Street, since High Street has the lowest best-so-far score. However, once the next stretch is revealed to us, and we calculate the new best-so-far routes, we see that Main Street has the best score in stretch 8; moreover, Main Street in stretch 8 points back to Main Street in stretch 7. Thus what seems like the best choice for stretch 7 at the time turns out not to be the best choice for stretch 7, given what happens subsequently. In this way the dynamic programming model gives a nice account of an important phenomenon in music perception: the fact that we sometimes revise our initial analysis of a segment based on what happens later. We will return to this phenomenon—which I call "revision"—in chapter 8.

In a recent article, Desain, Honing, vanThienen, and Windsor (1998) argue that, whenever a computational system is proposed in cognitive science, it is important to be clear about which aspects of the system purport to describe cognition, and which aspects are simply details of implementation. As explained earlier, the "model" in the current case is really the preference rule systems themselves. There are probably many ways that a preference rule system could be implemented; the dynamic programming approach proposed here is just one possibility. However, the dynamic programming scheme is not without psychological interest. It provides a computationally efficient way of implementing preference rule systems—to my knowledge, the only one that has been proposed. If humans really do use preference rule systems, any efficient computational strategy for realizing them deserves serious consideration as a possible hypothesis about cognition. The dynamic programming approach also provides an efficient way of realizing a preference rule system in a "left-to-right" fashion, so that at each point, the system has a preferred analysis of everything heard so far—analogous to the process of real-time listening to music. And, finally, dynamic programming provides an elegant way of describing the "revision" phenomenon, where an initial analysis is revised based on what happens afterwards. I know of no experimental evidence pertaining to the psychological reality of the dynamic programming technique; but for all these reasons, the possibility that it plays a role in cognition seems well worth exploring.[10]